

On the propriety of the posterior of hierarchical linear mixed models with flexible random effects distributions

F. J. Rubio*

Abstract

The use of improper priors in the context of Bayesian hierarchical linear mixed models has been studied under the assumption of normality of the random effects. We study the propriety of the posterior under more flexible distributional assumptions and general improper prior structures.

keywords: Half-Cauchy prior; Improper priors; Skew-normal; Two-piece normal.

1 Introduction

Hierarchical linear mixed models (LMM) are often used to account for parameter variation across groups of observations. The general formulation of this type of models is given by the model equation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{y} is an $n \times 1$ vector of data, $\boldsymbol{\beta}$ is a $p \times 1$ vector of *fixed effects*, \mathbf{u} is a $q \times 1$ vector of *random effects*, \mathbf{X} and \mathbf{Z} are known design matrices of dimension $n \times p$ and $n \times q$, respectively, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of residual errors. We assume that $\text{rank}(\mathbf{X}) = p$ and $n > p$ throughout. The random vectors $\boldsymbol{\varepsilon}$ and \mathbf{u} are typically assumed to be normally distributed. Given that the normality assumption can be restrictive in practice, alternative distributional assumptions have been explored, such as finite mixtures of normal distributions (Zhang and Davidian, 2001), scale mixtures of skew-normal distributions (Lachos et al., 2010), and Bayesian nonparametric approaches (Dunson, 2010). However, in a Bayesian context with improper priors, only the model with normal assumptions has been studied. In this line, Hobert and Casella (1996) analysed the following hierarchical structure:

$$\begin{aligned} \boldsymbol{\varepsilon}|\sigma_0 &\sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n) \\ \mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0 &\sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_0^2 \mathbf{I}_n), \\ \mathbf{u}|\sigma_1, \dots, \sigma_r &\sim N_q(\mathbf{0}, \mathbf{D}), \end{aligned} \quad (2)$$

*UNIVERSITY OF WARWICK, DEPARTMENT OF STATISTICS, COVENTRY, CV4 7AL, UK. E-mail: Francisco.Rubio@warwick.ac.uk

with an improper prior structure as follows:

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r) \propto \prod_{i=0}^r \frac{1}{\sigma_i^{2a_i+1}}, \quad (3)$$

where $\mathbf{u} = (\mathbf{u}_1^\top, \dots, \mathbf{u}_r^\top)^\top$, \mathbf{u}_i is $q_i \times 1$, $\sum_{i=1}^r q_i = q$, $\mathbf{D} = \oplus_{i=1}^r \sigma_i^2 \mathbf{I}_{q_i}$, $i = 1, \dots, r$, and $N_d(\mathbf{m}, \mathbf{S})$ denotes a d -variate normal distribution with mean \mathbf{m} and covariance matrix \mathbf{S} . The r subvectors of \mathbf{u} correspond to the r different random factors in the experiment. These assumptions imply that the random effects are assumed to be independent. This independence assumption restricts the applicability of this sort of models, however, it covers some models of interest in meta-analysis (Hobert and Casella, 1996; Tan and Hobert, 2009). Hobert and Casella (1996) obtained necessary and sufficient conditions for the propriety of the corresponding posterior distribution. This kind of improper priors may be of interest in the lack of strong prior information, although their use often represents a controversial topic. In particular, the choice $a_0 = 0$ and $a_1 = \dots = a_r = -1/2$ is referred to as the *standard diffuse prior* (Tan and Hobert, 2009). Sun et al. (2001) studied an extended improper prior structure which allows for assigning either proper or improper priors on the parameters of the distributions of the residual errors and the random effects as follows:

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r) \propto \prod_{i=0}^r \frac{1}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2). \quad (4)$$

This prior corresponds to assigning an inverse gamma distribution to the σ_i^2 's when $a_i > 0$ and $b_i > 0$, and it contains the prior structure (2) for $b_0 = \dots = b_r = 0$ (note that here we are presenting the priors on $(\sigma_0, \sigma_1, \dots, \sigma_r)$ rather than the equivalent priors on $(\sigma_0^2, \sigma_1^2, \dots, \sigma_r^2)$ presented in Hobert and Casella, 1996 and Sun et al., 2001).

In this paper, we explore extensions of the hierarchical structure (2)–(4) by considering more flexible random effect distributions. In Section 2, we present an extension of this model for the case when the random effects are distributed according to a two-piece normal distribution. We show that the corresponding posterior is proper essentially under the conditions presented in Sun et al. (2001) for the model with normal assumptions (2)–(4). In Section 3, we characterise a rich class of parametric distributions, which are obtained by adding a shape parameter to the normal distribution, that also preserve the existence of the posterior distribution when they are used for modelling the random effects in (1). In Section 4, we introduce an alternative improper prior structure with proper heavy-tailed priors on the scale parameters $(\sigma_1, \dots, \sigma_r)$. We conclude with a discussion on the implementation of this sort of models as well as possible extensions of this work.

2 Two-piece normal random effects

A random variable $U \in \mathbb{R}$ is said to be distributed according to a two-piece normal distribution, denoted $U \sim \text{TPN}(\mu, \sigma, \gamma)$, if its density function can be written as (Arellano-Valle et al., 2005):

$$s(u|\mu, \sigma, \gamma) = \frac{2}{\sigma[a(\gamma) + b(\gamma)]} \left[\phi\left(\frac{u - \mu}{\sigma b(\gamma)}\right) I(u < \mu) + \phi\left(\frac{u - \mu}{\sigma a(\gamma)}\right) I(u \geq \mu) \right], \quad (5)$$

where ϕ denotes the standard normal density function, I denotes the indicator function, and $\{a(\cdot), b(\cdot)\}$ are positive differentiable functions. This density is continuous, unimodal, with mode at $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_+$, and skewness parameter $\gamma \in \Gamma \subset \mathbb{R}$. It coincides with the normal density when $a(\gamma) = b(\gamma)$, and it is asymmetric for $a(\gamma) \neq b(\gamma)$ while keeping the normal tails in each direction. The most common choices for $a(\cdot)$ and $b(\cdot)$ correspond to the *inverse scale factors* parameterisation $\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$, $\gamma \in \mathbb{R}_+$ (Fernández and Steel, 1998), and the ϵ -skew parameterisation $\{a(\gamma), b(\gamma)\} = \{1 - \gamma, 1 + \gamma\}$, $\gamma \in (-1, 1)$ (Mudholkar and Hutson, 2000).

Consider now the following extension of the hierarchical structure (2),

$$\begin{aligned} \varepsilon|\sigma_0 &\sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n) \\ \mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0 &\sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_0^2 \mathbf{I}_n), \\ u_{ik}|\sigma_i, \gamma_i &\stackrel{ind.}{\sim} \text{TPN}(0, \sigma_i, \gamma_i), \quad k = 1, \dots, q_i, \quad i = 1, \dots, r, \end{aligned} \quad (6)$$

with prior structure

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r, \gamma_1, \dots, \gamma_r) \propto \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\gamma_i)}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2), \quad (7)$$

where $\pi(\gamma_i)$ are proper priors, and $\mathbf{u}_i = (u_{i1}, \dots, u_{iq_i})^\top$. The following result presents necessary and sufficient conditions for the propriety of the corresponding posterior.

Theorem 1 Consider the LMM (1) with hierarchical structure (6)-(7), and denote

$t = \text{rank}\{(\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{Z}\}$. Suppose that there exist constants $m, M > 0$ such that $h(\gamma) = \min\{a(\gamma), b(\gamma)\} < m$ and $H(\gamma) = a(\gamma) + b(\gamma) > M$. Consider also the following conditions:

(a) For $i = 1, \dots, r$ either $a_i < b_i = 0$ or $b_i > 0$,

(b1) $q_i + 2a_i > 0$,

(b2) $q_i + 2a_i > q - t$ for all $i = 1, \dots, r$,

(c1) $n - p + 2 \sum_{i=0}^r a_i > 0$,

(c2) $n - p + 2a_0 + 2 \sum_{i=1}^r \min(0, a_i) > 0$,

(d) $\prod_{i=1}^r \int_{\Gamma} \frac{h(\gamma_i)^{q_i+2a_i}}{[a(\gamma_i) + b(\gamma_i)]^{q_i}} \pi(\gamma_i) d\gamma_i < \infty$,

(e) $\prod_{i=1}^r \int_{\Gamma} H(\gamma_i)^{2a_i} \pi(\gamma_i) d\gamma_i < \infty$.

There are two cases:

1. If $t = q$ or if $r = 1$, then the conditions (a), (b2), (c1), and (d) are necessary, while (a), (b2), (c2), and (e) are sufficient for the propriety of the posterior for almost any sample.
2. If $t < q$ and $r > 1$, then the conditions (a), (b1), (c1), and (d) are necessary, while (a), (b2), (c2), and (e) are sufficient for the propriety of the posterior for almost any sample.

Proof. See appendix.

Conditions (a)–(c2) are exactly the same as those presented in Theorem 2 from Sun et al. (2001) for the propriety of the posterior associated to the hierarchical structure (2)–(4), while the conditions (d)–(e) are specific to the skewness parameter of the two-piece normal extension proposed here. Conditions (a)–(c2) refer to the need for the presence of certain amount of repeated measurements, while conditions (d)–(e) are trivially satisfied for some parameterisations, such as the ϵ –skew parameterisation where $H(\gamma) \equiv 2$. The ϵ –skew and the inverse scale factors parameterisations satisfy the boundedness conditions in Theorem 1. Note also that the bounds m and M do not appear in conditions (a)–(e) but they are essential for the proof of the result. The function $H(\gamma)$ in condition (e) can be replaced by $H(\gamma) = \max\{a(\gamma), b(\gamma)\}$, which might facilitate checking this condition under some parameterisations.

A discussion on the meaning of the phrase “for almost any sample” in Theorem 1 seems appropriate. The propriety results presented in Theorem 2 from Sun et al. (2001) are satisfied for the case when $2b_0 + \text{SSE} > 0$, where SSE represents the sum of square errors of model (1). This condition, as discussed by Sun et al. (2001), is satisfied with probability one (this is, for almost any sample) given that the distributions involved in this model are continuous. Given that the proof of Theorem 1 is based on that of Theorem 2 from Sun et al. (2001), this (mild) condition is inherited by our proposed extension.

The elicitation of the proper priors $\pi(\gamma)$ can be made following the strategy proposed in Rubio and Steel (2014). They propose assigning a uniform prior on a measure of skewness which is an injective function of the parameter γ . This induces a proper prior on γ which reflects vague prior information on a quantity of interest. For the ϵ –skew parameterisation this strategy simply leads to a uniform prior on $\gamma \in (-1, 1)$.

3 Sufficient conditions for other extensions

A natural question is whether or not there are other distributions with shape parameters that can be used for modelling the random effects in (1) while preserving the existence of the posterior distribution. In order to answer this question, we consider a general extension of the normal distribution obtained by using the representation proposed in Ferreira and Steel (2006). Ferreira and Steel (2006) show that the transformed density function

$$s(u|\mu, \sigma, \lambda) = \frac{1}{\sigma} p \left[\Phi \left(\frac{u - \mu}{\sigma} \right) \middle| \lambda \right] \phi \left(\frac{u - \mu}{\sigma} \right), \quad (8)$$

where $p : [0, 1] \rightarrow \mathbb{R}_+$ is a continuous density function with shape parameter $\lambda \in \Lambda \subseteq \mathbb{R}$, and Φ denotes the standard normal distribution function, is also a density function. Moreover, for any continuous density s with support on \mathbb{R} , there exists a density function p such that the relation (8) holds. The density s in (8) is symmetric if and only if p is symmetric around $1/2$, and it coincides with the normal density when p is the uniform density function. Throughout we denote by $U \sim \text{FSN}(\mu, \sigma, \lambda; p)$ the representation (8) of the distribution of an absolutely continuous random variable U (FSN is used here as an acronym of the **Ferreira-Steel** representation of transformations of the **N**ormal distribution). The following result provides sufficient conditions on p for the existence of the posterior of model (1) with an improper prior structure and non-normal assumptions on the random effects.

Theorem 2 *Consider the LMM (1) with the following hierarchical structure:*

$$\begin{aligned} \varepsilon|\sigma_0 &\sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n) \\ \mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0 &\sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_0^2 \mathbf{I}_n), \\ u_{ik}|\sigma_i, \lambda_i &\stackrel{\text{ind.}}{\sim} \text{FSN}(0, \sigma_i, \lambda_i; p), \quad k = 1, \dots, q_i, \quad i = 1, \dots, r, \end{aligned} \quad (9)$$

with prior structure:

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r, \lambda_1, \dots, \lambda_r) \propto \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\lambda_i)}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2). \quad (10)$$

Suppose that the priors $\pi(\lambda_i)$, $i = 1, \dots, r$, are proper and that the density p in (8) is upper bounded. Then, the conditions (a), (b2), and (c2) in Theorem 1 are sufficient for the propriety of the corresponding posterior distribution for almost any sample.

Proof. See appendix.

This result implies that the use of skew-normal random effects (Azzalini, 1985, whose representation is obtained for the choice $p[\Phi(u)|\lambda] = 2\Phi(\lambda u)$, $\lambda \in \mathbb{R}$) in the hierarchical structure (9) leads to a proper posterior. This result also includes other symmetric (non-normal) distributions for a symmetric (with respect to $1/2$) and bounded choice of p . Other distributions for which the corresponding p is bounded are discussed in Murillo and Rubio (2013). The density p associated to the representation (8) of the two-piece normal density (5) is not upper bounded (Murillo and Rubio, 2013). This indicates that Theorem 1 is not a consequence of Theorem 2, which justifies their separate study. It also implies that the boundedness of the density p is not a necessary condition for the existence of the posterior.

4 An alternative weakly informative prior structure

The use of improper priors for the scale parameters in hierarchical models is certainly a debatable topic. However, the use of certain proper priors in this context has also foster some discussions. In particular, the inverse gamma distribution, which appears in the structure (4), has been shown to be influential on the shape

of the posterior distribution (Gelman, 2006). This is an undesirable property if one is interested on using weakly informative priors (which is usually the aim when using jointly improper priors). An alternative proper prior for scale parameters that has gained a lot of attention is the half-Cauchy prior. Polson and Scott (2012) show that this prior induces a posterior distribution with good frequentist properties. The following result presents conditions for the propriety of the posterior distribution induced by an improper prior structure with half-Cauchy priors on the scale parameters.

Corollary 1 *Consider the LMM (1) with hierarchical structure (6) and prior structure*

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r, \gamma_1, \dots, \gamma_r) \propto \frac{1}{\sigma_0^{2a_0+1}} \prod_{i=1}^r \frac{\pi(\gamma_i)}{1 + \frac{\sigma_i^2}{s_i^2}}, \quad (11)$$

where $\pi(\gamma_i)$ are proper priors, and $s_i > 0$, $i = 1, \dots, r$. Let $(X : Z)$ denote the entire design matrix. Then, the following conditions are sufficient, for almost any sample, for the propriety of the posterior distribution under any parameterisation $\{a(\gamma), b(\gamma)\}$:

- (a) $a_0 \geq 0$,
- (b) $\text{rank}(X : Z) < n$.

Proof. See appendix.

It is worth emphasising that if we use a Dirac delta prior for γ_i concentrated at a value γ_o such that $a(\gamma_o) = b(\gamma_o)$, we obtain a hierarchical model with normal residual errors and normal random effects with an alternative prior structure to that in (4). This new prior structure assigns half-Cauchy priors on the scale parameters $(\sigma_1, \dots, \sigma_r)$ which, as mentioned before, has been argued to be less influential on the posterior distribution than the inverse gamma prior in (4). This feature may be appealing to practitioners in contexts with little prior information.

The prior structure in (11) involves half-Cauchy priors with arbitrary positive scale parameters s_i . This can be useful to conduct sensitivity analyses on the choice of these hyperparameters.

Theorem 1 from Fernández et al. (1997) can be used to construct other flexible hierarchical models as follows. By using this result, we can assign an *arbitrary* random effects distribution $\mathbf{u}|\boldsymbol{\theta} \sim F(\cdot|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents the parameter vector of the distribution F , for the LMM (1) with normal residual errors $\varepsilon|\sigma_0 \sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n)$ and prior structure:

$$\pi(\boldsymbol{\beta}, \sigma_0, \boldsymbol{\theta}) \propto \frac{\pi(\boldsymbol{\theta})}{\sigma_0^{2a_0+1}},$$

where $\pi(\boldsymbol{\theta})$ is any proper prior. Conditions (a) and (b) in Corollary 1 are sufficient for the propriety of the corresponding posterior distribution. This covers the class of random effects distributions studied in Section 3, finite mixtures of normals, scale mixtures of normals, and skewed scale mixtures of normals with proper priors on the corresponding parameters.

5 Discussion

The main message of the paper is that non-normal distributions can also be used to model the random effects in hierarchical linear mixed models as well as other improper prior structures without the need for additional onerous conditions for the propriety of the posterior distribution. This is useful to construct models that are robust to departures of normality of the random effects.

Given that the proposed hierarchical structures (6) and (9) contain the model with normal random effects as a particular case (for certain choices of p), the appropriateness of the normality assumption can be evaluated via Bayes factors since only the common parameters have improper priors. In particular, the use of the Savage–Dickey density ratio can be easily implemented to do so, as long as we are able to obtain a posterior sample from the corresponding parameters.

The use of non-normal distributional assumptions on the random effects in the context of LMM usually complicates their implementation. A general approach for sampling from the posterior distribution of the proposed hierarchical structures consists of using a Metropolis within Gibbs algorithm (Roberts and Rosenthal, 2009). However, this can be computationally demanding in moderate or high dimensions. The development of Gibbs samplers for the hierarchical structures (6)–(9), as well as convergence analyses of these, represent interesting research directions. In the context of LMM with normal random effects these ideas have been studied in Hobert and Casella (1996) and Román and Hobert (2012).

Further research includes the study of LMM with flexible distributions for the residual errors.

The case of multivariate scale mixture of normals and improper priors: a warning case?

Another type of extension that can be of interest in practice consists of using distributions with heavier tails than those of the normal distribution for modelling the random effects. A natural candidate for this choice is the family of multivariate scale mixtures of normals, which is a rich family that includes symmetric models with heavier tails than normal. Recall first that a p –variate, $p \geq 1$, scale mixture of normal distributions (denoted $\text{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta; H)$) is defined as

$$g(\mathbf{x}) = \int_0^\infty \frac{\tau^{p/2}}{\det(2\pi\boldsymbol{\Sigma})^{1/2}} \exp\left[-\frac{\tau}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] dH(\tau|\delta).$$

where $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix and H is a mixing distribution. For example, when $H(\tau|\delta)$ is a Gamma distribution with parameters $(\delta/2, \delta/2)$, we obtain the p –variate Student- t distribution with δ degrees of freedom.

Consider now the LMM (1) with the following hierarchical structure:

$$\begin{aligned} \boldsymbol{\varepsilon}|\sigma_0 &\sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n) \\ \mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0 &\sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_0^2 \mathbf{I}_n), \\ \mathbf{u}_i|\sigma_i &\stackrel{\text{ind.}}{\sim} \text{SMN}_{q_i}(\mathbf{0}, \sigma_i^2 \mathbf{I}_{q_i}, \delta; H), \quad i = 1, \dots, r, \\ \pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r, \delta) &\propto \pi(\delta) \prod_{i=0}^r \frac{1}{\sigma_i^{2a_i+1}}, \end{aligned} \tag{12}$$

where $\pi(\delta)$ is a proper prior on the shape parameter $\delta \in \Delta \subseteq \mathbb{R}$. It is possible to prove that for this model the marginal likelihood of the data can be written as follows:

$$\tilde{m}(\mathbf{y}) = m_N(\mathbf{y}) \int_{\mathbb{R}_+^r \times \Delta} \left[\prod_{i=1}^r \tau_i^{-a_i} dH(\tau_i|\delta) \right] \pi(\delta) d\delta,$$

where $m_N(\mathbf{y})$ is the marginal likelihood of the data with normal assumptions (2)-(3). This result indicates that $\tilde{m}(\mathbf{y})$ is proportional to the marginal likelihood of the data associated to the model with normal assumptions. Moreover, the proportionality constant depends only on the choice of the hyperparameters a_i , the mixing distribution H , and the prior $\pi(\delta)$. Therefore, conducting model selection using Bayes factors between models of the type (12) with different mixing distributions (*e.g.* against the model with normal random effects) is not driven by data (!). This result provides a warning on the use of multivariate scale mixtures of normals combined with improper priors in this context. Intuitively, this result indicates that the data do not contain enough information to distinguish between models of type (12).

Extensions with other generalised linear mixed models

In the context of generalised linear mixed models (GLMM) with improper priors, only models with normal random effects appear to have been studied (Chen et al., 2002; Sun and Lin, 2010). Some limitations induced by the assumption of normality of the random effects have been analysed in Litière et al. (2008). Given that the proof technique employed in Theorems 1 and 2 is based on obtaining upper and lower bounds for the random effects distributions in terms of the normal distribution, this idea can also be applied to other GLMM with improper priors. In particular, the one-way random effect probit model studied in Sun and Lin (2010):

$$\begin{aligned} y_{ik}|\mu, u_i &\sim \text{Bernoulli}[\Phi(\mu + u_i)], \quad k = 1, \dots, q_i, \quad i = 1, \dots, r, \\ u_i|\sigma &\stackrel{\text{ind.}}{\sim} \text{N}(0, \sigma^2), \\ \pi(\mu, \sigma) &\propto \frac{1}{\sigma^{2a_1+1}}, \end{aligned}$$

can be extended to the use of two-piece normal random effects as follows.

Remark 1 Consider the following one-way random effect probit model,

$$\begin{aligned} y_{ik}|\mu, u_i &\sim \text{Bernoulli}[\Phi(\mu + u_i)], \quad k = 1, \dots, q_i, \quad i = 1, \dots, r, \\ u_i|\sigma, \gamma &\stackrel{\text{ind.}}{\sim} \text{TPN}(0, \sigma, \gamma), \\ \pi(\mu, \sigma, \gamma) &\propto \frac{\pi(\gamma)}{\sigma^{2a_1+1}}, \end{aligned} \tag{13}$$

where $\pi(\gamma)$ is a proper prior. The posterior distribution of (μ, σ, γ) is proper under the following conditions:

- (i) for each i , $i = 1, 2, \dots, r_1$, ($2 \leq r_1 \leq r$), there is at least one success and one failure,

$$(ii) \quad -(r_1 - 1)/2 < a_1 < 0,$$

$$(iii) \quad \int_{\Gamma} H(\gamma)^{2a_1} \pi(\gamma) d\gamma < \infty.$$

Proof. The result follows by using the proof technique in Theorem 1 together with Theorem 2.1 from Sun and Lin (2010).

In a similar way, this model can be extended to the class of random effects distributions discussed in Section 3, which, in particular, allows for the use of skew-normal random effects.

Remark 2 Consider the following one-way random effect probit model,

$$\begin{aligned} y_{ik} | \mu, u_i &\sim \text{Bernoulli}[\Phi(\mu + u_i)], \quad k = 1, \dots, q_i, \quad i = 1, \dots, r, \\ u_i | \sigma, \lambda &\stackrel{\text{ind.}}{\sim} \text{FSN}(0, \sigma, \lambda; p), \\ \pi(\mu, \sigma, \lambda) &\propto \frac{\pi(\lambda)}{\sigma^{2a_1+1}}, \end{aligned} \tag{14}$$

where $\pi(\lambda)$ is a proper prior and the density p is upper bounded. The posterior distribution of (μ, σ, λ) is proper under conditions (i) and (ii) from Remark 1.

Proof. The result follows by using the proof technique in Theorem 2 together with Theorem 2.1 from Sun and Lin (2010).

Other types of GLMM with improper priors where the assumption of normality of the random effects can be relaxed by using the ideas in this paper are presented in Section 4 of Chen et al. (2002). The corresponding extensions and propriety results are omitted for the sake of space.

Acknowledgement

FJR gratefully acknowledges support from EPSRC grant EP/K007521/1. I thank an Associate Editor and a referee for helpful comments.

Appendix

Proof of Theorem 1

Let us denote $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_r)$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_r)$. The marginal density of the data is given by

$$\begin{aligned}
m(\mathbf{y}) &\propto \int f(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0) f(\mathbf{u}|\boldsymbol{\sigma}, \boldsymbol{\gamma}) \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\gamma_i)}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma} d\boldsymbol{\gamma} \\
&\propto \int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi\left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0}\right) \\
&\quad \times \prod_{i=1}^r \prod_{k=1}^{q_i} \frac{1}{\sigma_i} \left\{ \exp\left[-\frac{u_{ik}^2}{2\sigma_i^2 b(\gamma_i)^2}\right] I(u_{ik} < 0) + \exp\left[-\frac{u_{ik}^2}{2\sigma_i^2 a(\gamma_i)^2}\right] I(u_{ik} \geq 0) \right\} \\
&\quad \times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\gamma_i)}{\sigma_i^{2a_i+1} [a(\gamma_i) + b(\gamma_i)]^{q_i}} \exp(-b_i/\sigma_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma} d\boldsymbol{\gamma}.
\end{aligned}$$

Denote $h(\gamma) = \min\{a(\gamma), b(\gamma)\}$, then we can obtain a lower bound, up to a proportionality constant, for $m(\mathbf{y})$ as follows

$$\begin{aligned}
&\int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi\left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0}\right) \prod_{i=1}^r \prod_{k=1}^{q_i} \frac{1}{\sigma_i} \exp\left[-\frac{u_{ik}^2}{2\sigma_i^2 h(\gamma_i)^2}\right] \\
&\times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\gamma_i)}{\sigma_i^{2a_i+1} [a(\gamma_i) + b(\gamma_i)]^{q_i}} \exp(-b_i/\sigma_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma} d\boldsymbol{\gamma}.
\end{aligned}$$

By using the change of variable $\vartheta_i = \sigma_i h(\gamma_i)$ and the lower boundedness of $h(\cdot)$ we obtain the following lower bound for $m(\mathbf{y})$:

$$\begin{aligned}
&\int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi\left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0}\right) \prod_{i=1}^r \prod_{k=1}^{q_i} \frac{1}{\vartheta_i} \exp\left[-\frac{u_{ik}^2}{2\vartheta_i^2}\right] \\
&\times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{1}{\vartheta_i^{2a_i+1}} \exp(-b_i m^2/\vartheta_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\vartheta} \\
&\times \int_{\Gamma^r} \prod_{i=1}^r \frac{h(\gamma_i)^{q_i+2a_i}}{[a(\gamma_i) + b(\gamma_i)]^{q_i}} \pi(\gamma_i) d\boldsymbol{\gamma}. \tag{15}
\end{aligned}$$

Now, denote $H(\gamma) = a(\gamma) + b(\gamma)$, then we can obtain an upper bound, up to a proportionality constant, for $m(\mathbf{y})$ as follows

$$\begin{aligned}
&\int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi\left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0}\right) \prod_{i=1}^r \prod_{k=1}^{q_i} \frac{1}{\sigma_i} \exp\left[-\frac{u_{ik}^2}{2\sigma_i^2 H(\gamma_i)^2}\right] \\
&\times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\gamma_i)}{\sigma_i^{2a_i+1} H(\gamma_i)^{q_i}} \exp(-b_i/\sigma_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma} d\boldsymbol{\gamma}.
\end{aligned}$$

By using the change of variable $\vartheta_i = \sigma_i H(\gamma_i)$ we obtain the following upper bound

$$\begin{aligned}
& \int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi \left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0} \right) \prod_{i=1}^r \prod_{k=1}^{q_i} \frac{1}{\vartheta_i} \exp \left[-\frac{u_{ik}^2}{2\vartheta_i^2} \right] \\
& \times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{1}{\vartheta_i^{2a_i+1}} \exp(-b_i M^2/\vartheta_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\vartheta} \times \int_{\Gamma^r} \prod_{i=1}^r H(\gamma_i)^{2a_i} \pi(\gamma_i) d\boldsymbol{\gamma}. \quad (16)
\end{aligned}$$

From (15) we can identify the first factor as the marginal likelihood of the data under normal residual errors ε and normal random effects \mathbf{u} . From (16) we can also identify the first factor as the marginal likelihood of the data under normal residual errors ε and normal random effects \mathbf{u} . Then, by combining (15), (16) and Theorem 2 from Sun et al. (2001) we obtain necessary and sufficient conditions for the finiteness of $m(\mathbf{y})$.

Proof of Theorem 2

Let us denote $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_r)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)$. By using that p in (8) is upper bounded, the marginal density of the data can be upper bounded, up to a proportionality constant, after integrating $\boldsymbol{\lambda}$ as follows:

$$\begin{aligned}
m(\mathbf{y}) & \propto \int f(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0) f(\mathbf{u}|\boldsymbol{\sigma}, \boldsymbol{\lambda}) \frac{1}{\sigma_0^{b+1}} \prod_{i=1}^r \frac{\pi(\lambda_i)}{\sigma_i^{a_i+1}} d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma} d\boldsymbol{\lambda} \\
& \propto \int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi \left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0} \right) \times \prod_{i=1}^r \prod_{k=1}^{q_i} s(u_{ik}|0, \sigma_i, \lambda_i) \\
& \times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\lambda_i)}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma} d\boldsymbol{\lambda} \\
& \leq \int \frac{1}{\sigma_0^n} \prod_{j=1}^n \phi \left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta} - \mathbf{z}_j^\top \mathbf{u}}{\sigma_0} \right) \prod_{i=1}^r \prod_{k=1}^{q_i} \frac{1}{\sigma_i} \phi \left(\frac{u_{ik}}{\sigma_i} \right) \\
& \times \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{1}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2) d\mathbf{u} d\boldsymbol{\beta} d\sigma_0 d\boldsymbol{\sigma}. \quad (17)
\end{aligned}$$

Therefore, $m(\mathbf{y})$ is upper bounded, up to a proportionality constant, by the marginal likelihood associated to a model with normal residual errors ε and normal random effects \mathbf{u} . Consequently, the sufficient conditions for the finiteness of this upper bound are also sufficient for the finiteness of $m(\mathbf{y})$, which are obtained from Theorem 2 in Sun et al. (2001), and the result follows.

Proof of Corollary 1

The result follows by noting that the prior (11) assigns proper priors on the parameters of the distribution of the random effects. This propriety implies that the marginal distribution of the random effects, say $p(\mathbf{u})$, is proper. Under these conditions, Fernández et al. (1997) (Theorem 1) show that conditions (a) and (b) are sufficient for the existence of the corresponding posterior distribution for almost any sample. In this case, the phrase “for almost any sample” refers to all the vectors $\mathbf{y} \in \mathbb{R}^n$ such that there is no solution for the equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$.

References

- Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* 128: 427–443.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12: 171–178.
- Chen, M. H., Shao, Q. M., and Xu, D. (2002). Necessary and sufficient conditions on the propriety of posterior distributions for generalized linear mixed models. *Sankhya* 64: 57–85.
- Dunson, D. B. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics* (Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. Eds.), pp. 223–273. Cambridge University Press, Cambridge.
- Fernández, C., Osiewalski, J., and Steel, M. F. J. (1997). On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics* 79: 169–193.
- Fernández, C., and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–371.
- Ferreira, J. T. A. S., and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association* 101: 823–829.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis* 3: 515–534.
- Hobert, J. P., and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* 91: 1461–1473.
- Lachos, V. H., Ghosh, P., and Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* 20: 303–322.
- Litière, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine* 27: 3125–3144.
- Mudholkar, G. S., and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference* 83: 291–309.
- Murillo, A., and Rubio, F. J. (2013). A note on the infinite divisibility of a class of transformations of normal variables. *Brazilian Journal of Probability and Statistics* 27: 601–607.
- Polson, N., and Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* 7: 887–902.

- Roberts, G. O., and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* 18: 349–367.
- Román, J. C., and Hobert, J. P. (2012). Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors. *Annals of Statistics* 40: 2823–2849.
- Rubio, F. J., and Steel, M. F. J. (2014). Inference in two-piece location-scale models with Jeffreys priors (with discussion). *Bayesian Analysis* 9: 1–22.
- Sun, D. and Lin, X. (2010). A note on the existence of the posteriors for one-way random effect probit models. *Statistics & Probability Letters* 80: 57–62.
- Sun, D., Tsutakawa, R. K., and He, Z. (2001). Propriety of the posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica* 11: 77–95.
- Tan, A., and Hobert, J. P. (2009). Blocks Gibbs sampling for Bayesian random effects models with improper priors: convergence and regeneration. *Journal of Computational and Graphical Statistics* 18: 861–878.
- Zhang, D., and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* 57: 795–802.